Quantifying Crime Waves

Jameson Toole,[1,*] Nathan Eagle,[2,†] and Joshua B. Plotkin[3,‡]

[1]*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*
[2]*The Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, New Mexico 87501, USA*
[3]*Department of Biology and Program in Applied Mathematics and Computational Science, The University of Pennsylvania, Philadelphia, PA 19104, USA*

Using a longitudinal data set containing high resolution spatial and temporal data of crimes committed in the city of Philadelphia during 1999, we attempt to quantify correlations and relationships between crime events as well as locations throughout the city. We use a variety of techniques from autoregressive models, time-series analysis, and random matrix theory to deduce the correlation structure within the data. Results show that the largest signal in the data is most likely an artifact of police procedure and find the strongest correlation of crime rate to day of the week, with Sundays experiencing low crime levels, while Tuesday and Wednesday having the highest. We show this daily correlation appears stronger for certain types of crime (mainly drug related incidents) and that crime rates for all other offenses can't be distinguished from noise. In addition, we find no significant pattern of causation where in the presence of one type crime at a location predicts the presence of another type later in time.

I. INTRODUCTION

The term *crime wave* is often used to loosely characterize trends in crime rates. Many theoretical frameworks exist for attempting to explain causal influences of crime and its ebb and flow. Theories range from broken windows to seasonal weather, but the trouble with these explanations is that they are difficult to quantify. Implicit in the way crime is portrayed is the notion that crime that is here and now can effect crime that is there and then. While there has been much work done in studying aggregated crime statistics over long time scales (years and decades) or coarse grained spatial regions (countries, states, urban areas, rural areas, ect.), little work exists quantifying the spread of crime on a much more local level (i.e. city blocks, neighborhoods). Even the work which has been performed on a finer grained mesh focuses predominantly on the socio-economic study of crime; tracking small groups of criminals or correlating crime with various socio-economic indicators within neighborhoods. While this work is informative and sometimes predictive, it leaves room for a spatio-temporal analysis of crime which may more quantitatively describe local crime trends.

Choosing the correct framework to use when analyzing crime data is a difficult task given the lack of knowledge about the underlying causes of crime. Socioeconomic, psychological, and even environmental variables are all contributing factors in flow dynamics. When describing this flow throughout a city, it is necessary to describe both spatial and temporal patterns. Analysis methods must allow for temporal lag as well as irregular and flexible spatial boundaries. While an oversimplified model might treat crime as a simple diffusion problem, we must also be able to detect when crime on one side of the city affects crime on the opposite side, while leaving the middle unchanged (this could be due to difference is police resource allocation or geographic variables such as transportation systems).

To best accomodate the above criteria, a multi-step analysis method is proposed which combines multivariate autoregressive (MVAR) methods with auto and cross-correlation statistics. Results from random matrix theory (RMT) provide a benchmark to distinguish signal from noise in our system.

II. EXISTING TREATMENT OF CRIME

While the focus of this paper is not a deep socioeconomic theory of crime, it is useful to keep existing frameworks in mind as a basic tool to check validity. Original work in criminology dates back to 19th century Europe where statisticians began looking at aggregate crime statistics across regions of France and The Netherlands. These early studies used cumulative data in regions which were fixed by government census and were primarily concerned with correlating wealth distributions with crime. There were also studies done on the various districts of Paris, comparing various socioeconomic indicators such as education level, to aggregate crime statistics[1], [2], [3].

Modern American criminology began in Chicago at the end of the 19th century. As American cities were experiencing rapid growth and frequent change, crime remained a large problem for urban planners. For the first time researchers began painstakingly gathering statistics on crimes and other factors on the street block and neighborhood level, allowing them to locate "crime hot spots". With another shift in population dynamics during the post-WWII baby boom and with the birth of suburbs, many new mechanism for crime were studied. Aggregate

*jltoole@umich.edu
†nathan@mit.edu
‡jplotkin@sas.upenn.edu

statistics, however, are still widely used simply because they are easily obtainable. LaFree provides an excellent review of techniques used to explain crime trends from the WWII to present [4]. "Broken windows" and "social disorganization" theories, popularized by Wilson and Kelling's article in the March, 1982 edition of The Atlantic, proposed crime as a consequence of urban decay and lack of community ownership in neighborhoods [5]. Neglected areas not only attract criminals, as the neglect is a sign of lack of police presence, but the the lack of order can have damaging effects on community morals, further driving crime.

The disorder within a neighborhood may be physical, social, and may even be *percieved*, not reflecting actual trends in crime. Sampson *et. all.* [6] has produced a large body of work exploring the complex relationships between disorder and crime. While violent crime is linked to neighborhood disorder, it is unclear that correcting actual disorder has any affect on perceived disorder, which may be a by product of ingrained stereotypes. There has also been work done examining the relationship between community strength and interconnectedness and crime rates. It may be the case that a cohesive community is enough to enforce social order and reduce crimes rates [7], [8].

Also on the neighborhood level, studies have been done to correlate not high risk areas but high risk people, where individuals are tracked based on socioeconomic class and categorized by their propensity to turn to crime. Crime levels in neighborhoods are then predicted based on the population profile in the area. For example, Krivo *et. al* has found significant differences in the circumstances under which whites and black commit violent crimes [9], [10], [11].

In this research, however, we see a gap. Statistical methods have been used extensively to characterize large, aggregate data sets over long periods of time, while sociological studies have been performed on more micro levels. There is a need, however, for high resolution, local analysis of crime which may prove both insightful and predictive.

III. DATA

The data being analyzed consists of all rimes reported within the city of Philadelphia during the year 1999. The data set includes roughly 200,000 crimes reported at 37,000 unique locations across Philadelphia. Each crime entry includes information about the time, location, and nature of the event. Locations are geocoded with precise latitude and longitude coordinates allowing spatial resolution on a block to block level. Plotting all reported crimes, the street layout of the city is visible (FIG. 1).

Along with the locations of crimes, temporal data, down to the hour the crime was reported, is also available. This allows us to analyze criminal activity at time

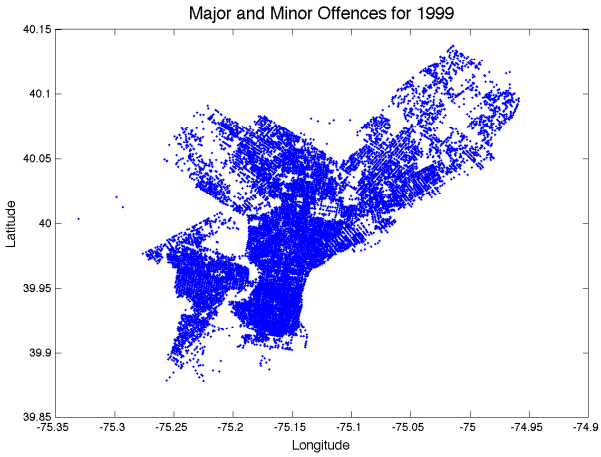


FIG. 1: A plot of all offenses, major and minor, throughout the city of Philadelphia during the year 1999. In total, 200,000 crimes were reported at 37,000 unique locations, each with high resolution spatial and temporal markers. Geographic features of the city, such as rivers, parks, bridges, ect., are immediately visible.

scales ranging from single days to an entire year. While the temporal density of crimes is too low to study on an hourly basis, daily crime levels are enough to reveal features such as holidays and seasonal trends (FIG. 2).

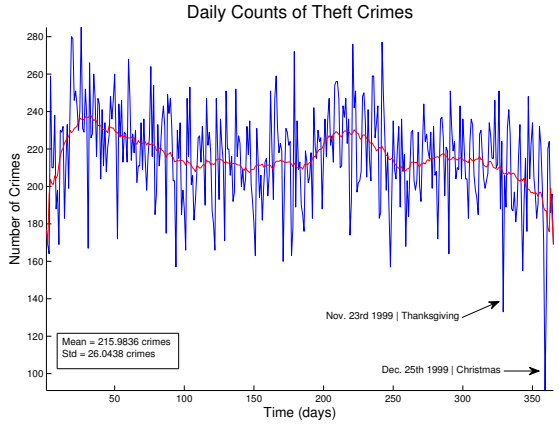


FIG. 2: A time series plot of theft crimes. Significant outliers can be identified as holidays such as Thanksgiving and Christmas. Smoothing the data (the red/central line) reveals seasonal trends where crime decreases in winter then peaking towards the end of summer during hot and humid weather.

Also coupled to this spatial and temporal data is a detailed description of the crime (burglary, automobile theft, assault with a particular weapon, ect.), making it possible to compare the flows of different crimes throughout the city. A detailed breakdown of the frequency with which different crimes occurred can be found in FIG. 3.

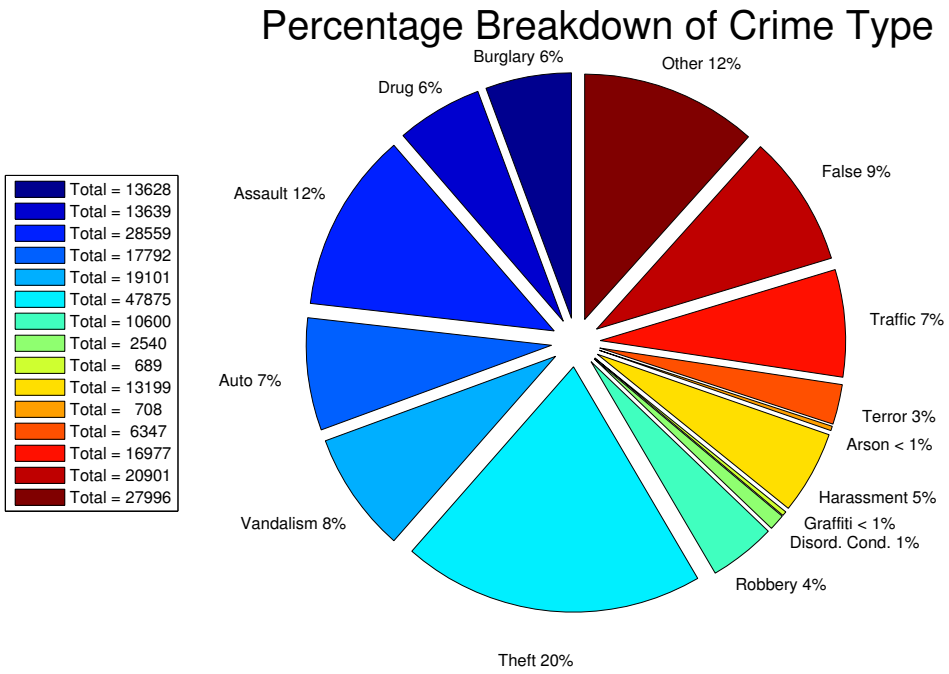


FIG. 3: A percentage breakdown of different crimes. Theft crimes, including burglary, robbery, and general theft make up the majority of offenses, while violent, drug related, and vandalism related incidents each account for significant portions. Note roughly 10% of all crimes are false reports.

In addition to broader, aggregate statistics, this high resolution data provides enough information to begin to look at crime flows on a micro level while still exploring similarities and differences between different types of criminal activity.

IV. METHODS

The ultimate goal of our analysis is to extract from the data information about crime flow across the city. It is our hope that any such analysis will capture any patterns or structure that may exist. Ideally, any analytic methods will capture diffusion in local space (from one neighborhood to an adjoining neighborhood), as well as diffusion across the city or other geographic boundaries. In addition to an irregular spatial structure, there must also be room for temporal lags (crime in one area effects crime in another area at a later time) and non-periodic correlation. Given these considerations, standard analytic methods such as diffusion models, Fourier analysis , wavelet analysis, and other analysis of low dimension provided little insight about the flow of crime.

For the main body of the article, we use auto and cross-correlation between time series allow us to directly detect the existence and magnitude of concurrent and delayed correlation. In addition to these relationships, the results generated from such analysis can also be used to inform multivariate autoregressive models (MVAR) as well as measure causality relations as defined by Granger Causality.

Correlation and causality measures are important to many disciplines and data. For the numerous bodies of literature that make use of time series data, there are equally man techniques to analyze them. We have attempted to adapt methods from neuroscience to finance markets to asses correlation relationships in our data. While not all of them yield significant results, in their failure we can learn something about our time series. For example, time series of neuron firing patterns and the networks that they form are well study by neuroscientists. These time series are characterized by spike train data in which large spikes appear periodically above a noisy base. Thus, their methods are tailored to correlate definite, periodic pulses in noisy time series. By examining which methods give reasonable results, we can characterize, in a broad sense, what type of data crime is. For additional information on a few other time series analysis techniques explored, see Appendix A.

A. Conditioning the Data

To use such techniques, we condition the data and construct time series as follows. Because different crimes likely exhibit different flow patterns and behavior, we divide crimes into categories based on the police reported description of each crime event. To ensure large enough

samples, six crime groupings were chosen (TABLE I). Making use of most of the data, these groupings account for nearly 75% of all crimes where nearly half the remaining crimes are false reports.

TABLE I: Categorical groupings of different crime types.

Category	Offenses Included	Number of Crimes (%)
All	all reported offenses	211606 (100%)
Automobile	Auto theft, major traffic violations	16005 (7.6%)
Theft	theft, burglary, robbery, auto theft	80290 (38%)
Violent	assault, rape, homicide, gun violence	29908 (14%)
Vandalism	vandalism, graffiti	17880 (8.5%)
Drug	possession, manufacturing, sale, DUI	11642 (5.5%)

Given all crimes from a certain category, we first lay a lattice grid structure over the city and then aggregate crimes onto the lattice point closest to their reported location. In most cases, our lattice structure consisted of 100 nodes, mirroring the spatial size of census tract districts. This grid size will allow for future correlation to census tract variables corresponding to individual neighborhoods, while also ensuring that crime count time series at each node will be sufficiently dense (ie. there a significant number of crimes reported at lattice point each day, with as few zeros as possible). For less reported crimes, only nodes with total crime counts above a certain threshold were kept, reducing the number of locations that could be correlated.

Temporally, crimes at each lattice point were binned at both daily and weekly intervals. Where possible, daily time series were constructed to allow more nodes to be analyzed. Unfortunately, the temporal resolution at which our analytical techniques were performed was hindered by the fact that as we decrease the size of each time interval, we must increase the number of intervals over which we look for lagged dependence. For example, if we suspect crime flows on the weekly scale, we must only account for 1 or 2 (week) lags, but our time series are only 52 periods long. If, however, we construct daily series, we now have 365 elements in each series, but must accommodate 7-14 (day) lags to capture the same dependence. As a technical note, the time series produced are stable and stationary, having roughly constant mean and variance over the year. Seasonal effects are only apparent for certain crime types and are weak at best. Unless otherwise stated, assume all time series to contain daily crime counts.

B. Correlation and Causality Measures

1. Granger Causality

Questions of flow inherently require knowledge of a temporal link. To understand how crime here and now influences crime at another location, later in time, there is need for a measure of causality across different time lags. A useful formalism of this problem was presented by Granger, an economist, to asses causal relations in econometric regressions. In general, we say that some variable Y *Granger Causes* X if information about past values of Y helps us make statistically better predictions of X, than just information about X alone [12].

It is also important to distinguish between *instantaneous causality* and *lagged causality*. The former describing the predictive power of *current* values of Y on X, while the latter of *past* value of Y on X. Instantaneous Granger Causilty makes more clear some problems inherent to any discussion of causality. While present values of X and Y may be correlated, inferring a causal relationship may be misleading, as there might be some other variable Z that influences both X and Y [13]. For more rigorous treatment of causality tests see work by Toda *et al.* [14] and Hacker *et al.* [15].

2. Pearson Correlation Coefficient

We begin our analysis by establishing a basic correlation matrix. In a purely statistical sense, the Pearson product-moment correlation coefficient is perhaps the most widely used and easily recognized of correlation measures. In words, we define the correlation coefficient, $\rho_{X,Y}$, between two random variables X and Y as the covariance, $cov(X,Y)$, divided by the product of their standard deviations, $\sigma_X\sigma_Y$. Mathematically this is can be written as

$$\rho_{X,Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sqrt{E[(X-\mu_X)^2]}\sqrt{E[(Y-\mu_Y)^2]}} = \frac{cov(X,Y)}{\sigma_X\sigma_Y} \tag{4.1}$$

If we assume X and Y are random variable vectors whose individual components are drawn from respective finite variance distributions. Geometrically, we can interpret the correlation coefficient as a measure of the cosine of the angle between the two vectors. As usual when dealing with sampled data, we rely on sample means and sample variance for calculation, but neither the formula nor the interpretation change significantly. It should also be noted that this measure says nothing of any temporal or lagged relationship between individual components of each vector, but simply states statistical correlation between underlying probability distributions. The matrix of correlation coefficients is a symmetric matrix with

whose diagonal is unity. It is also worth noting the eigenvalues of such a matrix will be real and positive.

3. (Lagged) Cross-Correlation/Autocorrelation

Cross-correlation and autocorrelation are measures applied more in signal processing and the physical sciences where periodicities are more prevalent or well defined signals appear buried under noise. Autocorrelation compares current values of a random process to past and future values of the same process, while cross-correlation compares two different random processes, possibly across lags. Most of the following derivation comes from Marple's text *Digital Spectral Analysis With Applications* [16].

Formally, the cross-correlation, r_{xy} between two random processes, X and Y, at times t_1 and t_2, is given by

$$r_{xy}[t_1, t_2] = E[x(t_1)y^*(t_2)] \tag{4.2}$$

where x and y are observations sampled from respective distributions (i.e time series). As a side note, a processes is called *wide-sense stationary* if the mean of the time series remains constant and two processes are called *jointly stationary* if the cross-correlation depends only on the time difference, $m = t_1 - t_2$. It may be the case that the random processes x and y have lagged cross-correlations. To sweep over a set of possible lags, we can define a *cross-correlation sequence* that is a series of cross-correlations over a range of time differences, m

$$r_{xy}[m] = E[x(t+m)y^*(t)] \tag{4.3}$$

Some properties of cross-correlation sequences that are helpful when analyzing data include

$$\begin{aligned} r_{xx}[0] &\geq |r_{xx}[m]| \\ r_{xx}[-m] &= r^*_{xx}[m] \\ r_{xx}[0]r_{yy}[0] &\geq |r_{xy}[m]|^2 \\ r_{xy}[-m] &= r^*_{yx}[m] \end{aligned}$$

In other words, we expect to see a large maximum in auto-correlation sequences around zero lag. Note also that $r_{xy}[-m] \neq r^*_{xy}[m]$. In general, autocorrelation series are symmetric around zero lag while cross-correlation are not. Furthermore, because elements of the cross-correlation series represent lags in time, it makes sense to transform the series to the frequency domain for further analysis. We define the power spectral density for two jointly stationary processes as a discrete-time Fourier transform (DTFT) of the cross-correlation sequence

$$P_{xy}(f) = T \sum_{m=-\infty}^{\infty} r_{xy}[m]e^{-i2\pi fmT} \tag{4.4}$$

where T is the sampling frequency. Note that for a signal which is pure white noise, all lagged auto-correlations are zero for non-zero lag ($r_{uu}[0] = \sigma_u^2$). Thus the auto-correlation sequence can be written as $r_{xx}[m] = \sigma_u^2\delta[m]$ where $\delta[m]$ is a discrete delta sequence. Transforming this to the frequency domain, we see that the DTFT of uncorrelated white noise is a constant, $P_{uu}(f) = T\sigma_u^2$. We use such measures to examine temporal links between crime in two different locations, or between different crime types at a single location. The cross-correlation at a specific lag can also give us insight into the speed at which a crime trend flows from one area to the next. If we construct two time series, the first ($\mathbf{y}_1$) in which counts are generated by a poisson process, the other ($\mathbf{y}_2$) a linear combination of past values the first plus a noise term, we expect to see a correlation series such as FIG. (4).

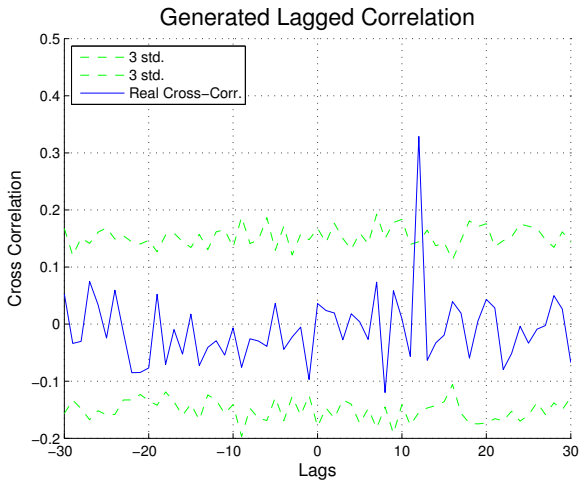


FIG. 4: The cross-correlation sequence two generated time series. Time series $y_1(t)$ was generated by a poisson random number generator while y_2 is a linear combination of values of $y_1(t-12)$ and white noise. The blue (solid) line represents the actual cross-correlation sequence while the green (dotted) lines represent 3σ intervals from a null model created by randomly permuting the time series and calculating the a distribution of cross-correlation series.

We can also relate these measures to Granger Causality as defined in the previous sections. Using the model presented in FIG. 4 as an example, we could say that variable $\mathbf{y}_1$ Granger causes $\mathbf{y}_2$ given the anti-symetric significant cross-correlation series.

4. Multivariate Autoregressive (MVAR) Models

Related to the calculation of cross-correlation sequences, are multivariate autoregressive models. Such models attempt to fit high dimensional output data to a number of input parameters that could be past val-

ues of output parameters or other, more abstract variables. This type of analysis appears across a wide variate of disciplines from economics to signal processing and there are a variety of well documents methods to apply to our data([17], [18]). Multivariate autoregressive models (MVAR) provide a fast and efficient way of correlating time series, including determining proper lag structure [19].

Unfortunately, however, these methods become unreliable given the large number of variables we are attempting to correlate. Given K time series, we wish to correlate time series $\mathbf{y}_i$ with each of the $K-1$ other time series accounting for p possible lags, we are attempting to fit the data to $K^2 \times p$ parameters. Given 100 time series and 30 lags, the least squares fit and statistical tests breakdown.

It is still possible, however, to use MVAR models to test basic hypotheses such as how crime yesterday (or last week) affects crime today as well as the effect that the day of the week has on reported crimes. The following, more detailed derivation of the standard MVAR model as applied to this crime data, comes from Lutkepohl 2007 [20] and Wooldridge 2008 [21].

Our most basic regression seeks to test the impact of crime one period before on crime now. Aggregating all data into a single time series whose values are the total number of crimes (in a given category) reported that day (or week) we regress the data onto the following equation.

$$\mathbf{y}_t = \beta \mathbf{y}_{t-1} + u_t \tag{4.5}$$

where u_t is a white noise term. Note that one crime type can also be regressed against past values of another crime type.

It is also possible to correlate crime counts at different locations with crime counts at other locations and times. Making reasonable assumptions such as stationarity and stability of our time series we build our model according to the standard vector autoregressive structure:

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \tag{4.6}$$

or,

$$\begin{bmatrix} y_{1,t} \\ \vdots \\ y_{K,t} \end{bmatrix} = \begin{bmatrix} \nu_{1,t} \\ \vdots \\ \nu_{K,t} \end{bmatrix} + \begin{bmatrix} a_{11,1} & \cdots & a_{1K,1} \\ \vdots & \ddots & \vdots \\ a_{K1,1} & \cdots & a_{KK,1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ \vdots \\ y_{K,t-1} \end{bmatrix} + \cdots + \begin{bmatrix} a_{11,p} & \cdots & a_{1K,p} \\ \vdots & \ddots & \vdots \\ a_{K1,p} & \cdots & a_{KK,p} \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ \vdots \\ y_{K,t-p} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ \vdots \\ u_{K,t} \end{bmatrix}$$

Included in the MVAR model is a fixed intercept ν to control for non-zero means among time series, as well as a *white noise* term. The noise variable u_t can be interpreted as fit residuals. Additionally, the elements, $a_{ij,n}; 1 \leq n \leq p$, of each coefficient matrix A_n can be qualitatively though of as the effect on $y_{i,t}$ due to a change in the value of $y_{j,t-n}$. Again, however, this method quickly becomes unreliable as the number of parameters grows large when many nodes and lags are incorporated. With very long time series, such methods may become more practical. For a more detailed description of the MVAR model including its close relation to auto and cross-correlation see Appendix B.

Given that our data set includes crimes which police responded to (from either 911 calls or police initiated actions) we can look for trends which may be the result of policing practice. For example, we may wish look for day dependent cycles where arrests decrease on Sunday's when less officers are on duty or increases on the Friday nights due to alcohol related incidents. To build a regression model that tests this, we make use of indicator variables for the day of the week,

$$\mathbf{y}_t = \beta \mathbf{y}_{t-1} + \sum_{k=1}^{7} \alpha_k \mathbf{I}_k(t) + u_t \tag{4.7}$$

where $\mathbf{I}_k$ is a vector length $T-1$ whose entries are 1 if entry (day) t is the kth day of the week and 0 otherwise.

5. *Random Matrix Theory (RMT)*

Many structural analysis techniques such as principal component analysis and singular value decomposition attempt to identify major driving forces in systems using the correlation and causation measures discussed in sections above. While the many techniques outlined above are efficient and effective ways to establish correlation and causation, it is important to distinguish to between random correlation and a true connection. To establish a sufficient null hypothesis we turn to random matrix theory, which provides analytic benchmarks for distinguishing signal from noise. These techniques have been used with much success in the analysis of data sets from financial (stocks, equities, ect.) to climate and weather. ([22],[23], [24], [25])

To obtain such distributions, we consider two related groups of matrices, Gaussian and Wishart. A matrix has a Gaussian distribution if each element is drawn from a standard normal. This is a convenient distribution to focus on because other distributions can be transformed into the standard normal simply and efficiently by subtracting the mean and dividing by the standard deviation. A Wishart matrix, $\mathbf{W}$, is then formed by matrix multiplication of a Gaussian matrix, $\mathbf{X}$, and its transpose, $\mathbf{W} = \mathbf{X}\mathbf{X}^T$ [26],[25]. It is important to note that the formulation of the Wishart matrix mirrors exactly the calculation of the cross-correlation matrix for zero lag.

Given that we have constructed random matrices as described above, we wish to find limits and distribution properties of the eigenvalues of such matrices that may

be used in the factor analysis or non-random systems. Asymptotic distributions are generally obtained by letting both K and T go to infinity [26], [27],[25].

We examine the preceding theory as it pertains to the patterns and structure we wish to find given our data. We begin by considering raw data in the form of a $K \times T$ matrix, $\mathbf{Y}$, where K is the number of time series we wish to correlate and T is the length if each time series. In order for the theory to be applicable, it is necessary to transform our time series to a random variables with zero mean and unite variance. The standard correlation matrix $\mathbf{C}$, not including any time lags, is constructed by matrix multiplication

$$\mathbf{C} = \frac{1}{T}\mathbf{Y}\mathbf{Y}^T \tag{4.8}$$

where ^{T} is a regular matrix transposition and the elements of $\mathbf{C}$ are given by $C_{ij} = \sum_{t=1}^{T} y_i(t)y_j(t)$. Thus eqn. (4.8) is just a matrix formulation of the Pearson Correlation Coefficient from section V. A. which is a special case of our cross-correlation measure. It is also of note that the correlation matrix $\mathbf{C}$ can be compared to analytical results that characterize Wishart matrices.

The derivation of the distribution of eigenvalues of a random Wishart matrix can be found in [26], [27],[25]. The formulations used in this analysis have been successfully applied to other data such as climate, weather, and financial in [23], [28], [25]. Defining the eigenvalue density, $\rho(\lambda)$ as the number of eigenvalues below λ and given a correlation (Wishart) matrix formed from completely random variables drawn from a standard normal distribution, the eigenvalue density distribution as K and T go to infinity, is given by the Marcenko-Pastur Law

$$\rho(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda} \tag{4.9}$$

where $Q = T/K \geq 1$ and $\lambda_{min}^{max} = 1 + 1/Q \pm 2\sqrt{1/Q}$ [23].

In other words, given the correlation matrix of completely random time series, we can expect the eigenvalue density of that matrix to follow the above distribution. Thus, we can use this result as a null hypothesis for data which is cannot be distinguished from noise. If we find eigenvalues significantly above or below these predicted thresholds, we can conclude there is a signal buried in the data, and further, perform factor analysis with confidence. For example, Laloux *et. al.* show that, while there is much noise in correlation matrices for various stocks and financial assets, there is still information in such noisy signals. The largest eigenvalue and (corresponding eigenvector) in the case of financial data is identified with the market as a hole, having equally weight components.

In addition to zero lag correlation matrices, analytical distributions for lagged correlations have also been put fourth by Mayya *et. al.* [29], [24]. If we want to study lagged correlations some additional manipulation of the data is necessary. We want to compare each time series to a time-shifted version of each other time series. To do this we define the matrix of unshifted time series, $\mathbf{Y}(0)$, and a matrix of shifted by p lags, $\mathbf{Y}(p)$. As a technical note, the shifted time series are padded with zeros so that the dimensions match in the matrix multiplication, while contributing nothing to the cross-correlation. We can then construct a time lagged cross-correlation matrix, $\mathbf{C}(p) = \mathbf{Y}(0)\mathbf{Y}^T(p)/T$, where the elements C_{ij} is the normalized lagged cross-corrlation between time series i and j.

Our analytical predictions, however, depend on symmetric matrices, but lagged cross-correlation is not a symmetric measure. We can construct a symmetric lagged cross-correlation matrix by averaging the lagged cross-correlation between i and j with the lagged cross-correlation between j and i. Mathematically we write this symmetric lagged cross-correlation matrix, $\mathbf{C}^S(p)$ as [29]

$$\mathbf{C}^S(p) = \frac{\mathbf{Y}(0)\mathbf{Y}^T(p) + \mathbf{Y}(p)\mathbf{Y}^T(0)}{2T} \tag{4.10}$$

Mayya *et. al.* have derived the eigenvalue density distribution for the above symmetric lagged cross-correlation matrix for zero mean and σ^2 variance time seres. The distribution is calculated by solving numerically a fourth order algebraic equation, keeping only the imaginary part [24].

$$\rho_p(\lambda) = \frac{1}{\pi}\lim_{\epsilon \longrightarrow 0} Im[G(\lambda - i\epsilon)] \tag{4.11}$$

where G is a complex function numerically solved for by

$$G^4 + 2\kappa G^3 + (\kappa^2 - \frac{Q^2}{\sigma^4})G^2 - 2\kappa\frac{Q^2}{\sigma^4}G + \frac{Q^2}{\sigma^4\lambda^2}(2Q-1) = 0 \tag{4.12}$$

and $\kappa = (Q-1)/\lambda$.

Finally, recent work by Thurner *et. al.* examines the spectra of random antisymmetric delayed correlation matrices [30]. Here we do not condition the data using eqn. (4.10) but instead leave the antisymmetric properties alone and look at both the real and imaginary parts of the eigenvalues and vectors of the delayed correlation matrix. Analytical derivation of such properties adds little to our discussion and follow similar lines to derivation in the symmetric case, thus we omit many theoretical results. To provide a sufficient null hypthosis we randomly permute each time series, calculate the delated correlation matrices, then create a distribution of eigenvalue spectra over 500 random permutations. From this distribution we calculate the maximum radius, $r_{max} = max(\sqrt{Re(\lambda_i)^2 + Im(\lambda_i)^2})$ for $1 \leq i \leq K$, of an eigenvalue. Given the maximum radius observed in random delayed correlation matrix spectra, we deem an

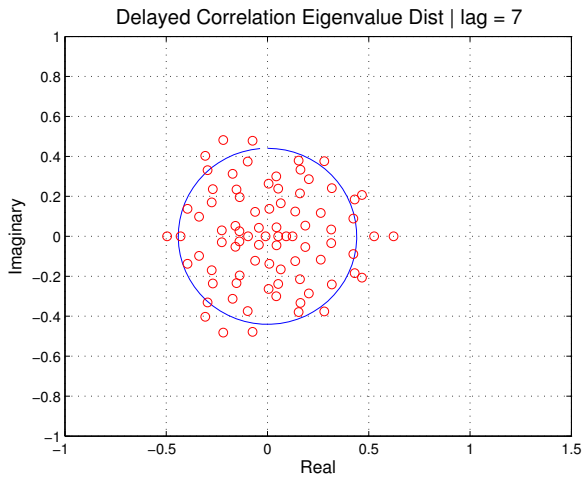


FIG. 5: A plot of the eigenvalue spectrum for the delayed correlation matrix violent crimes at a 7 day lag. The blue (solid) circle represents the maximum radius of an eigenvalue over 500 random spectra. Few eigenvalues lie outside this radius and those that do don't deviate far.

eigenvalue of the actual data significant if it falls outside of this circle in the complex plane (FIG. 5).

It is also informative to examine the components which make up the eigenvectors associated with the significant eigenvalues. For each normalized eigenvector, $\vec{\nu}_i$, we calculate its *inverse participation ratio* (IPR) to examine the component structure of the vector [31].

$$IPR(\vec{\nu}_i) = \sum_{j=1}^{K} |\nu_{ij}|^4 \tag{4.13}$$

A large IPR implies that only a few components contribute to the eigenvector, while a small IPR indicates participation of many components. It is also possible to determine clustering structure from such analysis. For example, in financial data the eigenvector corresponding to the large, "market", eigenvalue has a low IPR, identifying itself as a force which affects all stocks equally. Other eigenvectors, however, with larger IPRs have components which correspond to various sectors of the market [31]. For crime data, these components correspond to locations across the city so a cluster of eigenvector components would correspond to a cluster of neighborhoods.

Thus we now have a theoretical framework that can be used to differentiate between signal and noise in instantaneous and lagged cross-correlation matrices and are now ready to apply these methodologies to the crime data at hand.

V. RESULTS

Our first step in the analysis of crime data is to look for simultaneous correlations in the symmetric, zero-lag, correlation coefficient matrix whose entries are given by eqn. (4.1). We also examine the eigenvalue spectrum of this matrix and compare the distribution to the theoretical prediction from RMT and eqn. (4.9). Note that under our null hypothesis from RMT, our time-series are made up of white noise pulled from a standard normal distribution. Thus, we condition our data by taking each time series and subtracting the mean and dividing by the standard deviation for that particular time series, giving us zero mean and unit variance. Our analysis shows little visible correlation and an eigenvalue distribution that indicates a weak signal at best. The correlation structure shown in FIG. 6 is similar to those for all crime categories.

We first conclude that much of the correlation structure for drug related crimes cannot distinguished from noise and random correlations. The matrix spectra are in very close agreement with theoretical distributions for Gaussian noise. Similar results are obtained looking specifically at different crime categories. The lack of simultaneous correlation does not, however, tell us anything about any lagged correlation structure that may exist.

To examine lagged correlations, we employ the cross-correlation techniques discussed in previous sections. The cross-correlation series is computed for each node, comparing the time series at each location to every other location over a range of 30 lags. This allows us to examine periodicities and causal relationships between locations and crimes on the time scale of a month or less.

To test if significant correlations exist, we must first define a null hypothesis. We take two time series, $\mathbf{y}_1$ and $\mathbf{y}_2$, and compute the cross-correlation sequence as given by eqn. (4.3), for $0 \leq m \leq 30$. The values of the sequence at lag m, $r_{y_1y_2}[m]$, is the cross-correlaton between $\mathbf{y}_1$ and $\mathbf{y}_2$, where $\mathbf{y}_2$ has been shift m periods. This sequence represents the actual lagged correlation structure between two locations in the city. To test if there is any statistically significant structure, we randomly permute each time series and calculate the cross-correlation sequence for the randomized time series. We repeat this randomization 500 times, creating a distribution of random cross-correlation sequences for each pair of time series. From this distribution we can establish 3σ confidence intervals. We then say a correlation significant if the actual lagged cross-correlation coefficient, $r_{y_1y_2}[m]$, falls outside of the confidence interval established by our null distribution. Graphically, this procedure is represented by the generated data shown in FIG. 4.

Our results using this method are again a null result. Given K time series and looking over p lags, we generate $K^2 \times p$ cross-correlation coefficients. With all crime categories, we find that for all but drug related crimes, roughly 0.3% of these of these coefficients are deemed

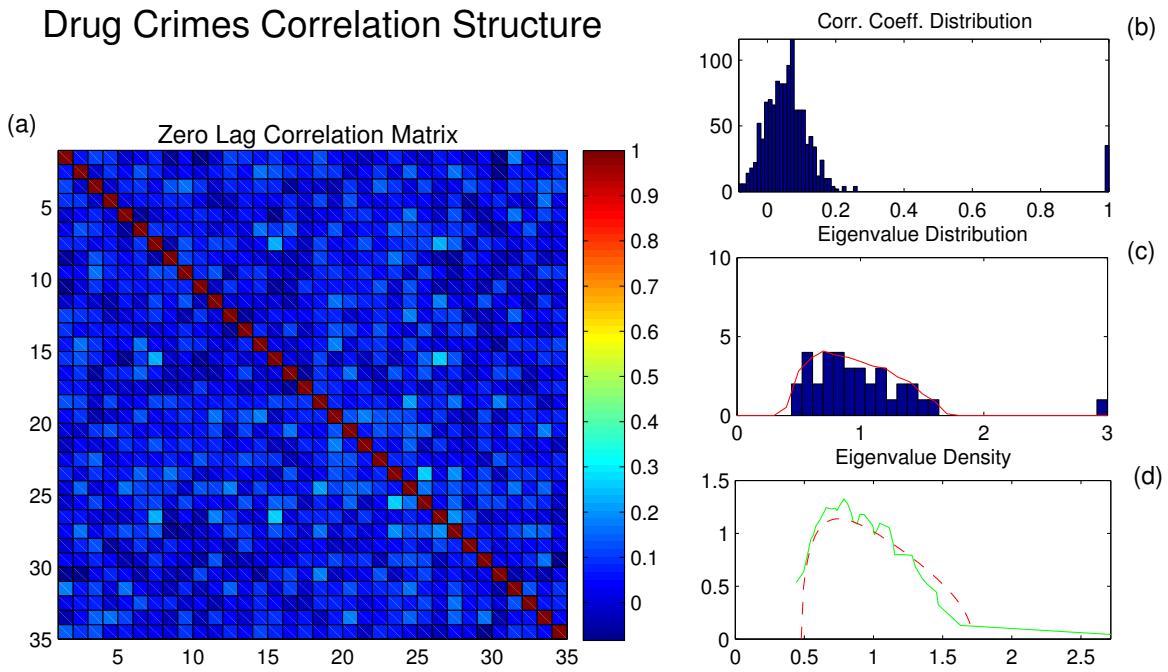


FIG. 6: (a) The zero lag correlation coefficient matrix for drug related crimes. No significant structure is immediately visible. (b) The distribution of correlation coefficients appears skewed slightly to the right. (c) The distribution of eigenvalues reveals one eigenvalue, magnitude 3, which can be differentiated from the noise indicated by the red (solid) line. (d) The green (solid) curve represents the eigenvalue density of the actual matrix spectra while the red (dashed) curve is the theoretical prediction from eqn. (4.9. From this analysis we find that there is very little significant zero lag correlation for drug related crimes.

significant above 3σ as defined in our null model. This is nearly exactly the percentage of random significant correlations one would expect to find, meaning we cannot reject the null hypothesis that the lagged cross-correlations were produced by means other than chance.

TABLE II: Significant lagged correlations by crime type. A total of 30 lags were accounted for each type. All but drug related crimes have a significant number of correlations

Category	Num. of Time Series (nodes)	Num. Sig. Correlations	Percent Significant (%)
Automobile	58	239	0.24%
Theft	118	936	0.22%
Violent	81	461	0.23%
Vandalism	65	350	0.28%
Drug	35	331	0.90%

As a final check for significant cross-correlation, we create a matrix whose entries are the sum of all significant lagged cross-correlation coefficients or 0 if no lagged cross-correlation exists between node i and node j for any lag. Visualizing this correlation matrix, we see no apparent structure which would indicate neighborhoods which are spatially close to eachother are correlated with respect to crime diffusion (FIG. 7).

Aside from the number and strength of significant lagged cross-correlations, we can also gain information from the distribution of these correlations in time. For instance, if the majority of significant correlations are found around a specific time lag, we may be able to quantify the time scale on which that particular type of crime propagates spatially. We also examine the spatial distance between two nodes given that they were significant correlated at some time lag. If, for example, crime diffuses locally in space, we expect nodes correlated at longer lags to have a greater spatial separation. We find little evidence, however, that either of these dynamics take place. In general, we find that the majority of the significant cross-correlations occur at zero lag (the exception being drug crimes that will be discussed later) and beyond this, no characteristic diffusion velocity is present. We also find conflicting evidence relating the temporal lag between two points and their spatial distance. Assuming all of our significant lagged cross-correlations are true connections and not random cor-

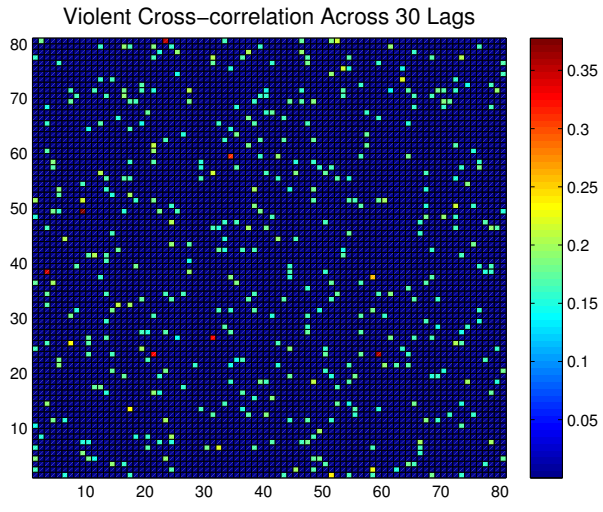


FIG. 7: The cumulative lagged cross-correlation matrix for violent crimes. Entries are the sum of significant lagged cross-correlation coefficients across 30 lags. If an entry is 0, this indicates no significant cross-correlation between node i and j over 30 days of lag. Because of the way nodes across the city were labeled, nodes j and $j+1$ are relatively close to each other spatially. Thus, if neighboring locations in the city were correlated in time, we should expect to see clusters of significant correlations, something we do not observe.

relations, we find that the automobile related crimes at nodes correlated at longer lags are slightly farther apart spatially, while for vandalism related crimes, the distance between two correlated nodes actually decreases as lag increases (FIG. 8).

While we perform the remaining analysis on all crime categories, we will focus heavily on the significant correlations found in drug related crimes. Turning our attention away from the cumulative significant cross-correlation matrices we examine the symmetric and antisymmetric lagged cross-correlation matrix for each individual lag and compare the spectra of such matrices to predictions from RMT.

We first examine the symmetric lagged cross-correllation matrix as defined by eqn. (4.10). The eigenvalues of this matrix of real, but may be positive or negative. We construct one symmetric lagged correlation matrix for each of our 30 lags. We can then look at the eigenvalue spectra for each delayed correlation matrix. Because of the small signal to noise ratio in our data, we focus on the largest eigenvalue of each of these matrices. This eigenvalue gives us a reasonable measure of the strength of the largest signal in the data. Plotting the magnitude of this eigenvalue for every lag, we see a strong pattern for drug related crimes, but hardly anything above noise for violent crimes (FIG 9).

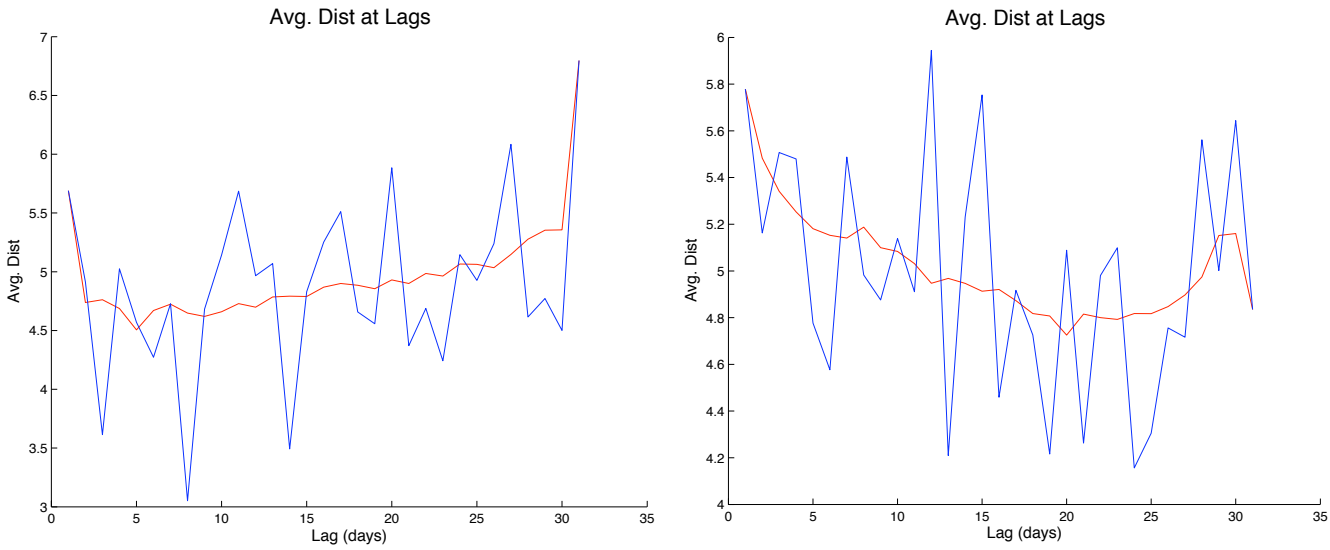
We will interpret these results shortly, but we first move one step further in our analysis by examining the antisymmetric delayed correlation matrices. Now that we are dealing with antisymmetric matrices, our eigenvalues and vectors may have both real and imaginary components. As mentioned in previous sections, the derivations of analytic distributions of eigenvalues are somewhat involved and add little to our application. Instead of computing them directly, we randomize our time series and compute random delayed correlation matrices. Repeating this process 500 times we take the maximum eigenvalue generated as the boundary distinguishing signal from noise. Again, we represent this maximum eigenvalue as a circle in the complex plane. Any eigenvalue that lies beyond this circle may be signal. Results here are similar to those from symmetric delayed correlation matrices. We find one large eigenvalue on the real axis for lags that are intervals of 7 days (1 week). For other lags, we find mostly noise (FIG. 10).

Finally, we would like to examine if any clustering can be done based on the significant eigenvectors of the delayed correlation matrix. We calculate the IPR for the eigenvectors of the delayed correlation matrix for a lag of 7, so as to capture as many significant eigenvalues as possible. Our results show that the eigenvector corresponding to the largest eigenvalue of this matrix has a low IPR and can thus be interpreted as our "market", with all of its components having near equal weight. Examining eigenvectors associated with the remaining significant eigenvalues, we find that they too have low IPRs, suggesting there is not any significant clusters or communities (FIG. 11). This agrees with our early results in which no clusters could be seen by examining cumulative and zero lag correlation matrices.

Before discussing the small signals we have found, we quickly summarize our results thus far. In general, zero lag correlation matrices reveal insignificant levels of correlation. Comparing the eigenvalue spectra of these matrices to theoretical distributions based on Wishart matrices reveals very little signal to distinguish from noise. Cross-correlation techniques accounting for lagged correlations also reveal few significant correlations and even then, the correlations do not display consistent trends. We find some signal, however, examining the eigenvalue spectra across lags, though this signal is weak for most crime types. Examining the spectra of even the most significant delayed correlation matrices does not reveal any community structure or causal relationship.

What, then, is driving the 7-day periodic cycle in significant correlations for drug related crimes? We first note that the signal does not drop in amplitude as the time lag gets greater. Presumably, if crime was truly spreading from one area to another, the importance of crime in the past to crime today would diminish as the past gets further and further away. This suggests we are not observing a flow of crime, but rather an artifact of reporting which is biasing our data. To examine this bias, we turn to simple autoregressive (AR) models.

The most basic AR model, relating crime today (this week) to crime yesterday (last week) was given by eqn.



(a) Automobile related crimes.

(b) Vandalism related crimes.

FIG. 8: Note that the y-axis distance measure is the distance across the lattice placed over the city and not standard distance units. Given that a significant correlation is found between two nodes for a certain lag, we calculate the spatial distance between these two nodes and compare it to the size of the time lag between the two time series. (a) For automobile related crimes there is a slight trend for two nodes that are further apart to have significant correlations for larger time lags. (b) For vandalism related crimes, the opposite trend appears. Given that these correlations may not be significant to begin with, we cannot conclude anything about the different speeds at which crimes may spread.

(4.5) in the previous sections. We begin by examining these models. In this case, we do not compare nodes across the city, but rather aggregate all crimes of a certain type across the entire city into one time series. Given the results from cross-correlation measures, it is unsurprising results based on this simple regression showed no significant or very weak relationship between crime today and crime yesterday (FIG. 12).

Given the steady weekly periodicity of the eigenvalues across lags, we next regress the data onto indicator variables corresponding to the day of the week. This model was given above by eqn. (4.7). With this model, we find significant correlation between crime counts and the day of the week. Regression coefficients and 95% confidence intervals can be found in TABLE III

Examining our regression coefficients we can immediately see significant results. Despite the relatively large confidence intervals, it is clear that Sunday and Monday have lower crime rates, while the middle of the week (Tuesday, Wednesday, and Thursday) see a steep increase in crime. Somewhat surprisingly, Friday's are not significantly correlated with crimes and Saturday even shows a slight decrease. One might expect that Friday and Saturday nights would have increased crime rates due to differences in weekend versus weekday activities.

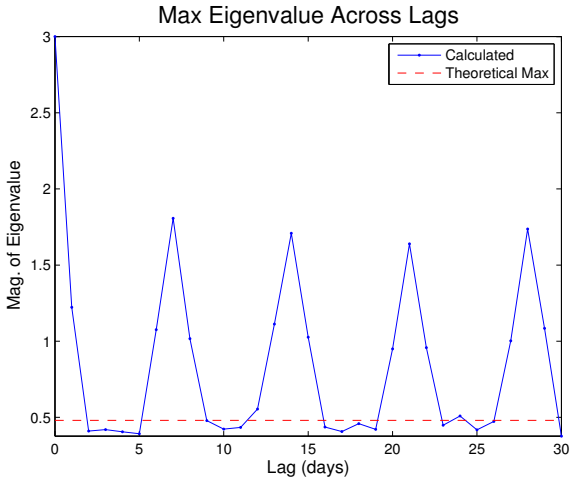
We now wish to see if this regressive model is enough to recreate the patterns found in the drug related crime data. We begin by creating a null model of time series which are constructed from our regression coefficients.

TABLE III: A simple autoregressive model accounting for crime yesterday as well indicator variables for the day of the week ($R^2 = .62$). It is worth noting that adding more time lagged terms does not increase the goodness of fit.

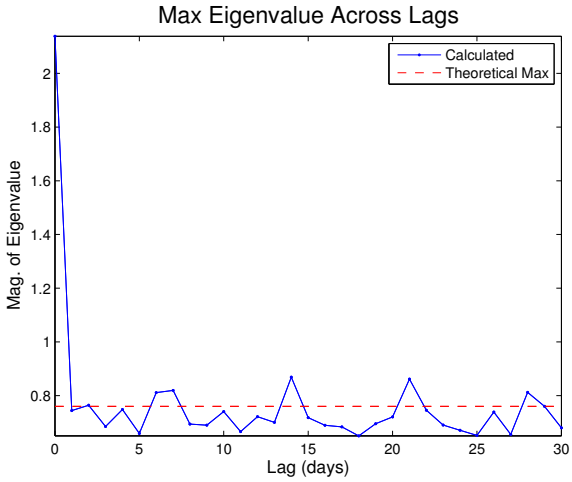
Input Variable	Coefficient	95% Confidence
Crime Yesterday	0.21	[0.11, 0.31]
Sunday	-1.1	[-1.3, -0.93]
Monday	-0.55	[-0.76, -0.34]
Tuesday	0.94	[0.75, 1.1]
Wednesday	0.75	[0.56, 0.93]
Thursday	0.47	[0.27, 0.66]
Friday	0.04	[-0.14, 0.22]
Saturday	-0.51	[-0.68, -0.34]

Crime counts for the first day are generated by a poisson process, then future counts are given by eqn. (4.7), the regression coefficients, and a noise term selected from a standard normal distribution. Performing all of the analysis outlined above, we match nearly exactly, the statistics for actual drug related crimes (FIG. 13).

In addition to zero-lag correlation measures, we also find our null model successfully reproduces the weekly spikes seen in the amplitude of the maximum eigenvalue of the symmetric lagged correlation matrices as shown in



(a) Drug related crimes.



(b) Violent crimes.

FIG. 9: We plot the maximum eigenvalue of the symmetric delayed correlation matrix for each of 30 lags. (a) For drug related crimes we see a very clear periodicity at a frequency of 7 days (1 week). (b) A much weaker pattern is seen for violent crimes, with the signal almost covered by noise. This suggests that there is some correlation for seven day periods. In each case, the red (dashed) line represents the maximum predicted eigenvalue from a random matrix as found by numerically solving for the distribution of eqn. (4.11).

FIG. 9a.

Our methods may also be used to look for relationships not between locations, but between crimes. For example, we may wish to ask if an increase in theft related crimes leads to violent crimes in the future. To look at these effects, we begin as before, making a time series of crime counts for each node in the lattice placed over the city. We create these time series for each crime category, thresholding to ensure significant sample data at each node. We then look for nodes that have enough

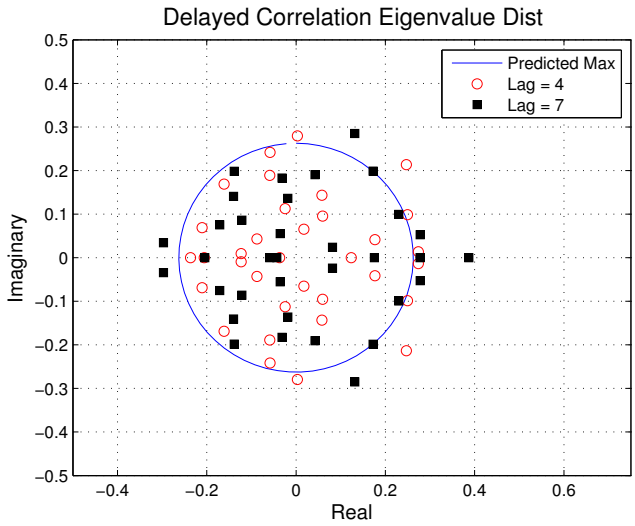


FIG. 10: The eigenvalue distributions for antisymmetric delayed correlation matrices for drug related crimes. The red (cicles) represent eigenvalues for a 4 period lag, while the black (squares) are eigenvalues of the 7 period lag matrix. The blue (solid) circle represents the maximum eigenvalue expected given randomized time series. In general there is a large eigenvalue on the real axis for lags which are multiples are 7 (1 week) while the distribution for other lagged matrices is much closer to pure noise.

crime at that location to use the time series for each crime type. From 128 nodes, we find 34 nodes that each have 5 time series, one for each crime category. We can then use all of the above techniques to examine any (lagged) correlation between different crimes.

We first compare two different crime types. For example we compare drug related crimes to violent crimes or thefts to automobile crimes. We use the latter as an example for the type of relationship our methods detect. We have included automobile thefts in both the "Thefts" category and the "Automobile" category. It is not surprising, then, we see significant correlation between the two crime types at exactly zero lag. To establish significance of values in the cross-correlation sequence between two crime types, we use the null model as before, randomizing each time series 500 times and creating a distribution from which we gain confidence intervals for actual data.

We would like to correlate two crime types at $N = 34$ nodes across, $p = 30$ lags (forwards and backwards) To visualize these correlations we create an $L \times N$ matrix where L is $(2p + 1)$. The rows of this matrix correspond to the cross correlation between the two crime types at a fixed lag across all 34 nodes. The columns of this matrix represent the cross-correlation sequence between the two crime types at each node. We threshold the elements of this matrix, setting them to zero if the correlation is deemed non-significant by our null hypothesis and, if

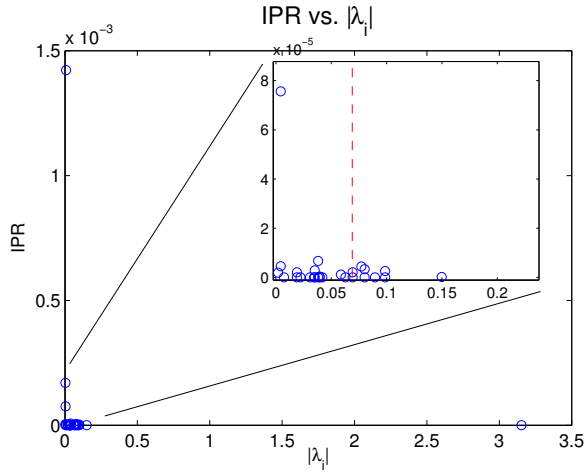


FIG. 11: Plotting the IPR for eigenvectors corresponding to the eigenvalues of the 7-period lag correlation matrix for drug related crimes, we find that the largest eigenvector corresponds to the "market" with a low IPR indicating equal weight among components. For other significant eigenvalues, we find low IPRs as well, indicating a lack of community structure. All eigenvalues to the left of the red (dashed) line cannot be distinguished from noise.

significance is established, setting an element equal to the normalized lagged cross-correlation coefficient. FIG. 15 shows this matrix for the comparison of automobile crimes to theft crimes at each location.

As was the case correlating different locations, we find that the number significant correlations for each of the 10 pairs (5 crime types comparing two at a time) is nearly exactly the number of random correlations to be expected above 3σ threshold.

For comparisons where there were a at least some correlation coefficients above 3σ significance, we can examine the location of these correlations. If, for instance, we find the majority of the correlations occur for negative lags, this tells us that the second crime type "causes" the first and vice versa. If a particular row in our matrix holds the majority of significant correlations, this indicates the time scale of flow. For automobile and theft crimes, this row corresponded to zero lag because we had included automobile crimes in theft crimes. In general, however, we find no significant correlations between different crime types at the same location, and furthermore, assuming those that are found are true connections, there is no definite pattern indicating the direction of the causal relationship or the speed of the flow.

VI. DISCUSSION

Analysis of the actual crime data and the success of our simple autoregressive model at recreating observed patterns suggests that the strongest signal in the data is simply the response of crime rates to the day of the week. Beyond this relationship, our methods reveal few significant signals. What, then, can we say about the flow of crime?

When stating a null result such as this it is important to qualify our statements. Our analysis can only answer very specific questions about crime. We have examined the relationship between crime levels one location to crime levels at another location for a specific spatial aggregation. Our analysis finds little significant correlation between different locations for any crime type. This result is somewhat surprising given the crime attractor and detractor theories that hypothesize criminals some neighborhoods should be sinks or sources for crime. In theory, our methods should allow us to draw a directed graph and this identify major attractors and detractors of crime. However, we find no significant correlations exist from which to draw this graph. Even from those correlations that exist, but may be random, we see no spatial patterns. That is, we do not find that neighborhoods which are close together form communities. Very little clustering exists with respect to correlations. This is not to say, however, that there is no correlation between neighborhoods when it comes to crime. Our analysis looks only at crime levels and does not take into account any socioeconomic or demographic data.

Perhaps a more robust result, is the lack of correlation between different crime types. Broken windows and social disorganization theories postulate that an influx of minor offenses such as graffiti and vandalism might lead to an increase of more serious crimes such as assaults or gun violence. We find no evidence of this in the data. Again though, this is not to say no relationship exists. We have ruled out interaction on time scales of up to 30 days, but it may be that these types of flows happen on the monthly or yearly time scale, thus it is not surprising that our analysis does not catch these correlations. The length of our time series, however, prohibits us from addressing this issue.

What we can say is that looking solely at interaction between crime levels in different neighborhoods or between different crime types on time scales of less than one month reveals little significant structure. This result may limit the power of predictive models that could be used by law enforcement to more efficiently allocate resources. There may still be merit, however, in predicting crime by coupling socioeconomic and demographic variables with neighborhoods, focusing more on at risk individuals than areas.

Given the amount and depth of criminology research which has already been done, it is surprising that we find so little structure in the data. One possible explanation is simply a sampling problem. There are well documented issues with using only police records as data because they are subject to misreporting or biases in the discretion of officers (i.e. there may be more than one way to record a crime). Aside from these issues, the largest sampling bias is that we only obtain data on the crimes which are

reported. Of all the crimes committed it is difficult to determine if those which are reported represent an unbiased sample of the total distribution. It may be the case that the crimes which are reported are simply sampled from the random part of the whole distribution. In other words, the majority of criminals that get caught (thus crimes that get reported) were caught because they are bad criminals or because of a chance passerby or some other fluke. It is reasonable to think that "good" criminals may be very organized, creating elaborate correlation structures between neighborhoods or types of crime, but these crimes do not appear in our data set because they go unreported.

It may also be the case that the majority of the crimes committed are crimes of opportunity. If the opportunity to steal a car or vandalize a building is the result of being in the right place at the right time, only by chance, it would make sense that we find little correlation between these chance opportunities. There is undoubtedly both planned and unplanned criminal activity, but it may be the case that the unplanned activity washes out any signal due to organized crime.

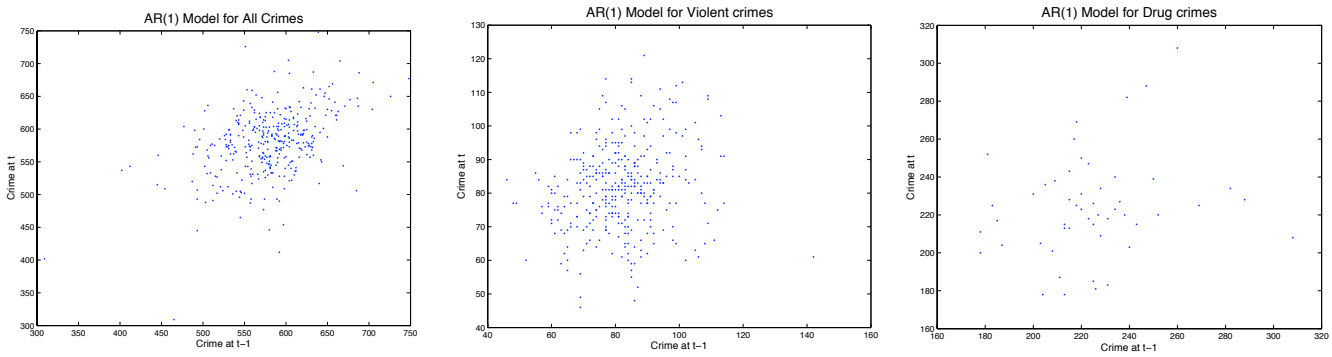
Finally, it is worth remembering what techniques can and cannot not be used. For example, we had difficulty using techniques to analyze spike train data, because criminal activity does not have the same structure spike train data. In contrast, methods used by researchers looking at time series of financial data faired better, suggesting that crime is more like the random walk of the stock market than neural patterns.

VII. CONCLUSION

Clearly there are circumstances under which our analysis fails. While we do not attempt to argue that crime is completely random, we provide convincing evidence that at small spatial and temporal scales, little correlation structure exists between crime levels at locations or between the levels of different types of crimes. These crime levels, of course, are subject to any sampling bias of the law enforcement system. It is possible that given more data and longer time series, patterns may emerge. Given this type of high resolution data, it may be more advantageous to develop better models based on socioeconomic variables, but this analysis is saved for another work.

VIII. ACKNOWLEDGMENTS

First and foremost I would like to thank Nathan Eagle for his time, ideas, patients, and support during this research. I would also like to thanks Joshua Plotkin for providing us with this data set. Finally, I would like to thank the SFI Community as a whole. To all those who provided me with ideas and guidance, I truly appreciate your time and effort. This research would not have been possible without your help.



(a) All Crimes (b) Violent Crimes (c) Drug Crimes

FIG. 12: Simple AR(1) Models of crime counts reveal little link between crime yesterday (or last week) and crime today (this week). (a) and (b) show regressions of daily rates, while (c) represents weekly counts.

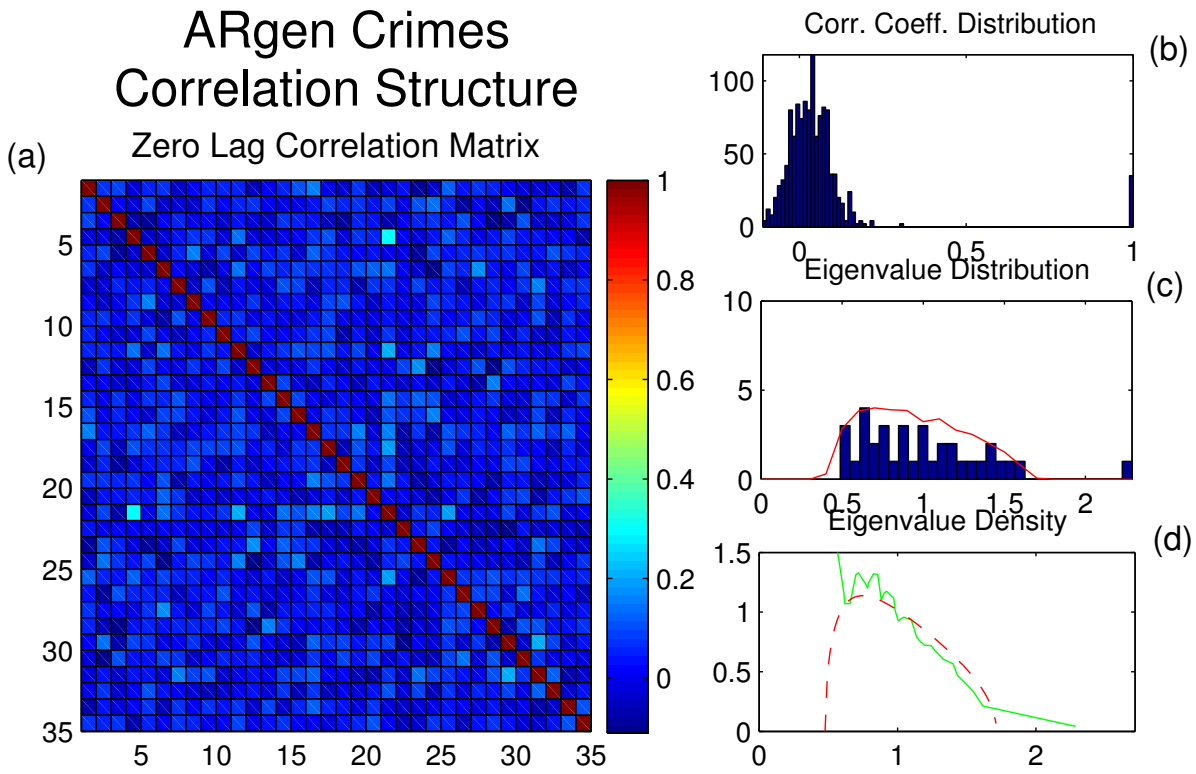


FIG. 13: The zero-lag correlation matrix and spectra for the null model generated from the regression coefficients of our day of the week drug crime model. These results are in very close aggrement with those from actual crime data shown in FIG. 6.

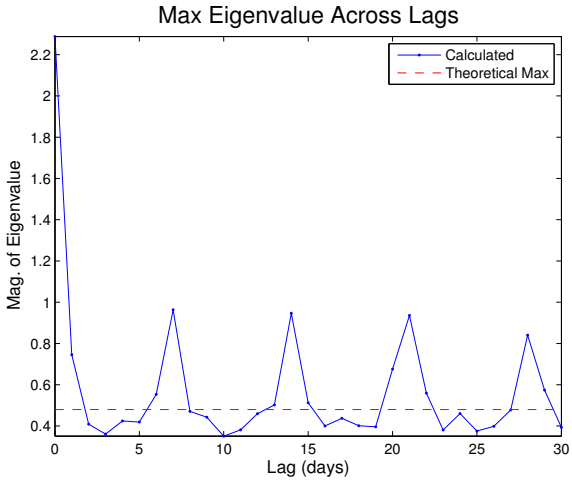


FIG. 14: The maximum eigenvalue of the symmetric delayed cross correlation matrix for each of 30 lags. The null model generated from regression coefficients captures the weekly periodicity of actual drug related crimes shown in FIG. 9a.

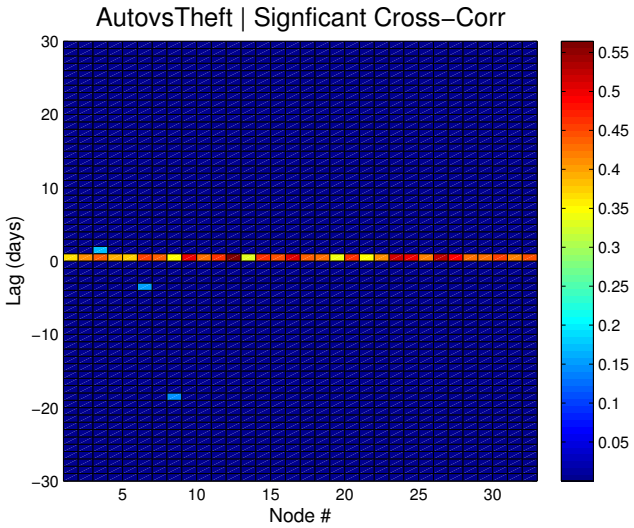


FIG. 15: A matrix displaying significant lagged cross correlations between automobile crimes and theft crimes. Because automobile crimes are counted in both categories, we find correlation at zero lag, but almost no other significant relationships.

Appendix A: Other Correlation and Causality Measures

1. Directed Transfer Function/Partial Directed Coherence

Given the coefficient matrices of an MVAR model described in section IV B 4, it is possible to make many inferences about causal relations between variables. From these relationships we can characterize directed flows within the system that can ultimately be used to create a directed network. As it turns out, inferring directional flows across a network given various time series data is a problem central to many questions in neuroscience. For example, the following techniques have been developed and applied successfully to the study of epileptic seizures, tracking brain activity as it flows through across a neural network. These methods, however, seem to work the best when there are relatively few nodes (20-50) and very long time series (most brain activity is sampled at 1000Hz for minutes at a time). Furthermore the time series represent spike train data, which is characterized by noise time series and large, detectable spikes corresponding to neuron firing. For a good comparison of there types of correlation techniques see Winterhalder 2005 [32]

a. Directed Transfer Function (DTF)

The directed transfer function, developed by Kaminski *et. al.* [33], seeks to determine causal influences among multichannel sources.

We can derive the directed transfer function from an multivariate VAR(p) model which compares K time series vectors across p lags. Given K time series $\mathbf{y}_1, \ldots, \mathbf{y}_K$. Restating the VAR(p) process

$$\mathbf{Y}(t) = \sum_{n=1}^{p} \mathbf{A}(n)\mathbf{Y}(t-n) + \mathbf{U}(t) \tag{A1}$$

Computing the parameters using any of the techniques described in previous sections, we obtain a number of series of coefficients, $A_{ij}(n), \text{for } 1 \leq n \leq p$. We then examine the spectrum of these coefficient series by transforming them into the frequency domain via an Fourier transform, $\mathbf{A}(f) = \mathbf{I} - \sum_{n=1}^{p} \mathbf{A}(n)e^{-i2\pi jn}$. We can now define the following relationship between time series in the frequency domain.

$$\mathbf{A}(f)\mathbf{Y}(f) = \mathbf{U}(f) \Rightarrow \mathbf{Y}(f) = \mathbf{A}^{-1}(f)\mathbf{U}(f) \tag{A2}$$

$$\Rightarrow \mathbf{H}^{-1}(f) = \mathbf{A}(f) \tag{A3}$$

where $\mathbf{H}$ is called the *transfer matrix*. Using this matrix we can characterize channel $j's$ influence on channel i as following noting that this influence is analogous to a measure of Granger causality.

$$\theta^2_{j\to i} = |H_{ij}(f)|^2 \tag{A4}$$

a normalized version of this is written as

$$\gamma^2_{ij}(f) = \frac{|H_{ij}(f)|^2}{\sum_{m=1}^{K} |H_{im}(f)|^2} \tag{A5}$$

In addition to these statistics, it is possible to characterize the magnitude of the total directed influence between variables i and j by summing up the series of VAR(p) coefficients, $\sum_{n=1}^{p} A^2_{ij}(n)$.

It is difficult to come up with analytical asymptotic distributions for DTF function, thus null models used to test significance have to be generated. One method proposed is to create significance statistics by randomize each time series then performing the algorithm again. Doing this many times give an ensemble of randomization to serve as null model. Further applications of the directed transfer function (DTF) can be found in [34].

b. Partial Directed Coherence (PDC)

An alternative measure to the DTF, claiming more precise structural analysis of signals, is *partial directed coherence (PDC)* as defined by Baccala *et. al* [35]. Similar to the DTF in that it analysis the spectra of coefficients estimated by a VAR(p) model, the PDC seeks to solve some of the problems of the DTF by measuring influence between two variables after the influence due to all other variables has been subtracted out. This addresses problems of a third variable contaminating causal relations between two others.

We must first introduce notions of *ordinary coherence*, which is defined using elements from the cross-spectral density matrix. From previous sections, the cross-spectral power density is defined by the cross-correlation series in eqn. (4.4). We then define the cross-spectral density matrix as

$$\mathbf{P}(f) = \begin{bmatrix} P_{11}(f) & P_{12}(f) & \cdots & P_{1K}(f) \\ P_{21}(f) & P_{22}(f) & \cdots & P_{2K}(f) \\ \vdots & \vdots & \ddots & \vdots \\ P_{K1}(f) & P_{K2}(f) & \cdots & P_{KK}(f) \end{bmatrix} \tag{A6}$$

From this matrix we define ordinary coherence as follows

$$C_{ij}(f) = \frac{|P_{ij}(f)|^2}{P_{ii}(f)P_{jj}(f)} \tag{A7}$$

Qualitatively, this measures the degree of simultaneous activity between channels i and j. To address the problem of directed coherence. We first factor the cross spectral density matrix as $\mathbf{S}(f) = \mathbf{H}(f)\Sigma\mathbf{H}^{\dagger}(f)$, where

the superscript † denotes the Hermitian transpose. We interpret $\mathbf{H}$ as a matrix of filters which act on the covariance matrix, Σ. From this factorization, a measure of directed corherence is defined as

$$\gamma_{ij}(f) = \frac{\sigma_{jj} H_{ij}(f)}{\sqrt{S_{ii}(f)}}, \; S_{ii}(f) = \sum_{j=1}^{K} \sigma_{jj}^2 |H_{ij}(f)|^2 \quad \text{(A8)}$$

The use of γ_{ij} here is not coincidental. When all σ_{jj} are set to 1, the DTF is recovered. As a result, all previous discussion of Granger causality in the sections about VAR models and the DTF are valid here. Unfortunately, it is often difficult to obtain consistent statistically significant results based on these measures. As such, *partial direct coherence (PDC)*, $|\kappa_{ij}(f)|^2$, is introduced using the following factorization

$$\kappa_{ij}(f) = \frac{\bar{\mathbf{a}}_\mathbf{i}^\dagger(f)\Sigma^{-1}\bar{\mathbf{a}}_\mathbf{j}(f)}{\sqrt{\bar{\mathbf{a}}_\mathbf{i}^\dagger(f)\Sigma^{-1}\bar{\mathbf{a}}_\mathbf{i}(\mathbf{f})\bar{\mathbf{a}}_\mathbf{j}^\dagger(f)\Sigma^{-1}\bar{\mathbf{a}}_\mathbf{j}(f)}} \quad \text{(A9)}$$

where $\bar{\mathbf{a}}_\mathbf{i}(f)$ is the ith column of the matrix $\bar{\mathbf{A}}(f) = \mathbf{I} - \mathbf{A}(f)$. From this we can define a *partial directed coherence factor*

$$\pi_{ij} \equiv \frac{\bar{A}_{ij}(f)}{\sqrt{\bar{\mathbf{a}}_\mathbf{j}^\dagger(f)\Sigma^{-1}\bar{\mathbf{a}}_\mathbf{j}(f)}} \quad \text{(A10)}$$

Relating this to Granger causality from section IV B 1, if we remove Σ from the denomenator, we effectively remove instantaneous Granger causality. Mathematically, this amounts to setting Σ to the identity matrix (a diagonal matrix with unit covariances). The partial directed coherence (PDC) is then defined as this quantity and is essentially a measure of influence from variable j to variable i.

$$\bar{\pi}_{ij} \equiv \frac{\bar{A}_{ij}(f)}{\sqrt{\bar{\mathbf{a}}_\mathbf{j}^\dagger(f)\bar{\mathbf{a}}_\mathbf{j}(f)}} \quad \text{(A11)}$$

Some properties of the PDC include, $0 \leq |\bar{\pi}_{ij}(f)|^2 \leq 1$ and $\sum_{i=1}^{K} |\bar{\pi}_{ij}(f)|^2 = 1$.

In general, the PDC normalizes influence with respect to the *receiving* channel, while the DTF normalizes with respect to the *sending* channel. Another subtle distinction is that $\bar{\pi}_{ii}$ reflects the direct effect of a channel's past on its current value, while the diagonal elements of the DTF reflects any power density that cannot be explained by any other channels.

As an illustrative example of the PDC (taken from [35]) we construct a correlated set of five time series and perform the above algorithm. The time series are constructed as follows

$$\begin{aligned}
x_1(t) &= 0.95\sqrt{2}x_1(t-1) - 0.9025x_1(t-2) + u_1(t) \\
x_2(t) &= 0.5x_1(t-2) + u_2(t) \\
x_3(t) &= -0.4x_1(t-3) + u_3(t) \\
x_4(t) &= -0.5x_1(t-2) + 0.25\sqrt{(}2)x_4(t-1) + 0.25\sqrt{2}x_5(t-1) + u_4(t \\
x_5(t) &= -0.25\sqrt{2}x_4(t-1) + 0.25\sqrt{2}x_5(t-1) + u_5(t)
\end{aligned}$$

Using a PDC computation algorithm implemented in MATLAB, we find the following (not the diagonal represents the power spectra)

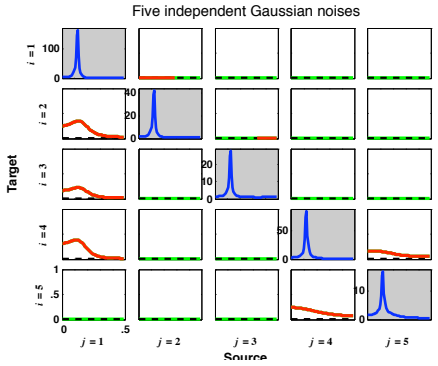


FIG. 16: A read out of the PDC computed on the above test system of correlated time series. Red curves are significant PDC measures, the diagonal is a power spectra [35].

Unfortunately, for our data, because we are not able to fit an MVAR model to more than one lag period, we are unable to generate the series of coefficients necessary for this analysis. Furthermore, it is difficult to address confidence measures to test for significance because the signal we are looking for is much different from that of spike train data. We do not see large peaks emerging from a noisy base in our data, thus it is unclear how to interpret the significance of results from these models when different types of data are used.

Appendix B: Further Treatment of MVAR Models

Beyond the basic formulation of the MVAR model, we may also use various techniques to infer causal relations from the correlation coefficients. For analysis purposes, it is useful to exploit the property that any VAR(p) model can be expressed as a VAR(1) model:

$$Y_t = \nu + \mathbf{A}Y_{t-1} + U_t \quad \text{(B1)}$$

where,

$$Y_t \equiv \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \nu \equiv \begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \text{(B2)}$$

$$\mathbf{A} \equiv \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_K & 0 & \cdots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{bmatrix}, U_t \equiv \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

This representation effectively transforms a VAR(p) model, comparing K time series over p lags, into a VAR(1) model comparing Kp dimensional vectors. Further more, we say a process such as Y_t is stable if the eigenvalues of $\mathbf{A}$ lie outside the unit circle, or

$$det(I_{Kp} - \mathbf{A}\lambda) \neq 0 \text{ for } |\lambda| \leq 1 \quad \text{(B3)}$$

Careful treatement of this stability criterion can be found in Lutkepohl 2001(Appendix C, Proposition C.9) [**?**], but for our purposes we use assume it holds in order to express Y_t as the sum of previous and current noise terms, $U_{t-n}, 0 \leq n \leq p$ and its average $\mu = E(Y_t) = (I_{Kp} - \mathbf{A})^{-1}\nu$. Thus we write Y_t as

$$Y_t = \mu + \sum_{i=0}^{\infty} \mathbf{A}^i U_{t-i} \quad \text{(B4)}$$

For a stable VAR(p) process, it is also possible to examine the autocovariance and autocorrelation structure of the model given. First we subtract μ and multiply both sides of eqn (4.6) by $(y_{t-n} - \mu)'$ (n is the number of time steps between the two observations), then take the expectation value of both sides (in other words computing the second moment of the VAR(p) process). Defining the second moment as $\Gamma(n) = E[(y_t - \mu)(y_{t-n} - \mu)']$. This moment is interpretadated as the covariance between random variables in period t and period $t - n$.

Computing the second moment of the VAR(p) process provides us with the *Yule-Walker equations* which allow the recursive calculation of $\Gamma(n)$. For $n \geq 0$

$$\begin{aligned} \Gamma_y(0) &= A_1\Gamma_y(1)' + \cdots + A_p\Gamma_y(p)' + \Sigma_u \\ \Gamma_y(n) &= A_1\Gamma_y(n-1) + \cdots + A_p\Gamma_y(h-p) \end{aligned} \quad \text{(B5)}$$

where Σ_u is a diagonal matrix whose elements are given by $E(u_t u_t')$.

Along with autocovariance, $\Gamma_y(n)$ can be used to calculate autocorrelations, a useful scale invariant measure. To determine the correlation between our vector now and our vector after n time steps, we first define a new diagonal matrix D such that the diagonal elements are the square roots of the diagonal elements of $\Gamma_y(0)$. In other words D is a diagonal matrix of the standard deviation of the random vector at each time step. Letting $\gamma_{i,j}(n)$ be the covariance between $y_{i,t}$ and $y_{j,t-n}$ (not this is also the i, jth element of $\Gamma_y(n)$) we define the autocorrelation between i and j over an n period lag as

$$\rho_{i,j}(n) = \frac{\gamma_{i,j}(n)}{\sqrt{\gamma_{i,i}(0)}\sqrt{\gamma_{j,j}(0)}} \quad \text{(B6)}$$

These autocorrelation and autocovariances are simply normalized versions of those mentioned in previous sections.

It is also worth noting here that Granger non-causality between two variables can be determined by examining the VAR(p) coefficients, $a_{ij,n}$ $1 \leq n \leq p$. We say $y_{i,t}$ is not Granger caused by $y_{j,t}$ if $a_{ij,n} = 0$ for all n. Again, however, even if a "causal" relationship is found between two variables, this does not rule out some third variable affecting of these variables. It should also be noted that if no causal relationship is found between two variables, it may be possible to transform the process such that the moments of the random variables remain the same, but a causal relationship is found. Finally, it may also be the case that the time resolution (months, weeks, days, ect.) influences which causality relations are found.

c. *Moving Average (MA) Representation*

To better analyze structural properties of autoregressive models, it is often useful to reformulate a model as a *moving average (MA) representation* of the the VAR process. To transform our VAR(1) model into a MA representation we start by examining eqn. (**??**). Note we recover the equation for the current random variable vector y_t by multiplying both sides of (B1)by matrix $J \equiv [I_K : 0 : \cdots : 0]$, a $K \times Kp$ matrix. Alternatively, with clever multiplication by J, we define a new $K \times K$ matrix $\Phi_i \equiv J\mathbf{A}J'$

$$y_t = JY_t = J\mu + \sum_{i=0}^{\infty} J\mathbf{A}J' = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i} \quad \text{(B7)}$$

A recursive relation for Φ_i is derived with the introduction of the *lag operator*, L, which simply moves the time index back one period such that $Ly_t = y_{t-1}$. We can re-write our VAR(p) model using this operator

$$y_t = \nu + (A_1 L + \cdots + A_p L^p)y_t + u_t \quad \text{(B8)}$$

Defining operators $A(L) \equiv I_K - A_1 L - \cdots - A_p L^p$ and $\Phi(L) = \sum_{i=0}^{\infty} L^i$ such that $\Phi(L)A(L) = I_K$ it is clear that $\Phi(L) = A(L)^{-1}$. Though we omit the details, this inversion relation is dependent on the stability condition (**??**). We are now able to define a recursion relation for Φ_i without resolving the VAR model.

$$I_K = \Phi(L)A(L) = (\Phi_0 + \Phi_1 L + \Phi_2 L^2 + \cdots) \cdot (I_K - A_1 L - \cdots - A_p L^p)$$

collecting powers of L,

$$I_K L^0 = \Phi_0 L^0 + (\Phi_0 - \Phi_1 A_1)L + (\Phi_0 - \Phi_1 A_1 - \Phi_2 A_2)L^2 + \cdots + (\Phi_i - \sum_{j=1}^{i} \Phi_{i-j} A_j)L^i + \cdots$$

now equating powers of L we see,

$$\begin{aligned} I_K &= \Phi_0 \\ 0 &= \Phi_0 - \Phi_1 A_1 \\ &\vdots \\ 0 &= \Phi_i - \sum_{j=1}^{i} \Phi_{i-j} A_j \\ &\vdots \end{aligned}$$

because we have assumed a only p period lag, $A_j = 0$ for $j > p$

$$\Rightarrow \Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j$$

Specifically for a VAR(1) process or a VAR(p) process modeled as VAR(1), it can be shown that $\Phi_0 = I_k, \ldots, \Phi_i = A_1^i, \ldots$. Thus we can easily compute coefficients of the moving average (MA) representation from the coefficients of the VAR(p) model.

To define test for Granger causality in the moving average representation, we start by writing the process in the form

$$y_t = \mu + \sum_{i=1}^{\infty} \Phi_i u_{t-i} \tag{B9}$$

where the system is characterized by the value of the white noise parameter introduced at each time step and the coefficient matrix Φ_i as defined above. In general, we can say that process $y_{1,t}$ is not Granger caused by $y_{2,t}$ if the element $\Phi_{12,i} = 0$for all$i = 1, 2, \ldots$. This condition is quiet similar to the Granger causality condition using the coefficients of the VAR model.

d. Structural Analysis

In the end, the point of using techniques such as VAR or MA processes is to gain some insight into to real causal relationships and influences between variables. Beyond a simple causal relationship, it may be interesting to know how a variable responds given a shock in another variable. Using the coefficients calculated in an MA model, we are able to analyze such dynamics. *Impulse response analysis* or *multiplier analysis* attempts to quantify the propagation of shocks through a linear system as described by the VAR and MA models.

To get an unobstructed view of how shocks in one variable effect another, it is important to build in some notion of *ceteribus peribus*. To model unit shocks in only one variable, we imagine a noise term at time $t = 0$, $u_0' = [u_{1,0}, \cdots, u_{K,0}] = [1, 0, \cdots, 0]$. Using this in our equation for the a general VAR process, eqn (4.6), we can see the system dynamics as follows

$$y_0 = \begin{bmatrix} y_{1,0} \\ \vdots \\ y_{K,0} \end{bmatrix} = u_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\Rightarrow y_1 = A_1 y_0 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{K,1} \end{bmatrix}$$

$$\Rightarrow y_i = A_1^i y_0$$

Similarly, you can place the shock in any element of u_0 such that $u_{j,0} = 1$ and the system will respond such that later values of y_t only depend on the jth column of A. Furthermore, since it has been shown that given a MA representation of a VAR process, $\Phi_i = A_1^j$ we can interpret elements of the Φ_i coefficient matrix as follows. Element $\phi_{jk,i}$ represents the response of variable, j, to a unit shock in variable, k, i periods ago. Thus plotting $\phi_{jk.i}$ as a function of i can give us some visual test for the strength of lagged correlation. Additionally, it is possible to plot cumulative impulses by summing up values of $\phi_{jk,i}$.

One problem with the above technique is that it assumes that shocks occur in only one variable at a time. This becomes problematic when variables are not independent, or many things are changing at once. Without drawing out the math, it is possible to free all other variables and test orthogonal impulses by forcing other variables to zero, thus shielding them from any shocks.

e. Link between Auto and Cross-Correlation to VAR and MA Processes

Given the similar goals of cross-correlation measures and VAR models, mainly to detect lagged correlations, it is no surprise that the two can be linked mathematically. In fact, auto and cross-correlation sequences can be used to calculate coefficients in both VAR and MA models (and can even be used in combinations of both). Marple [16] outlines criterion for selecting a model as follows.

To analyze spectrum with sharp peaks and few dead spots (deep nulls), a VAR model is appropriate, while

for spectra with no sharp peaks, but many nulls, an MA model might be more suited for analysis. It should also be a goal to minimize the number of parameters being fitted to allow for the best estimation statistics.

In general it is possible to model a VAR process using both VAR and MA representations. This new process is called an *autoregressive moving average* process (ARMA).

$$y[t] = -\sum_{n=1}^{p} a[n]y[t-n] + \sum_{n=0}^{q} b[n]y[t-n] \tag{B10}$$

where parameters $a[n]$ correspond to the VAR portion of the model and the, $b[n]$ to the MA portion. Here p is still the model order and q is the moving average parameter. If we set $b[0] = 1,\ b[n] = 0, \text{for} n > 0$ we recover a pure VAR model, while similarly setting $a[0] = 1,\ a[n] = 0, \text{for} n > 0$ recovers the MA model.

To calculate VAR and MA coefficients from a auto-correlation series, we first take our ARMA model (eqn (B10) and multiplying both sides by $y^*[t-m]$ and taking the expectation values giving

$$E[y[t]y^*[t-m]] = -\sum_{n=1}^{p} a[n]E[y[t-n]y^*[t-m]] + \sum_{n=0}^{q} b[n]E[u[t-n]y^*[t-m]] \tag{B11}$$

recognizing this as our definition of auto-correlation,

$$r_{xx}[m] = -\sum_{n=1}^{p} a[n]r_{xx}[m-n] + \sum_{n=0}^{q} b[n]r_{ux}[m-n] \tag{B12}$$

Because u_t represents a white noise vector we already have some information about the cross-correlations, $r_{ux}[m]$

$$r_{ux}[m] = \begin{cases} 0 & \text{for } i > 0 \\ \sigma_u^2 & \text{for } i = 0 \\ \sigma_u^2 h^*[-m] & \text{for } i < 0 \end{cases} \tag{B13}$$

We now substitute this into eqn. (B12) and recover a relationship between the autocorrelation sequence and ARMA parameters. While it is possible to also solve for MA parameters, this involves a more difficult algorithm. Thus we focus on the case where we are simply looking to build a VAR model (ie setting $b[0] = 1,\ b[n] = 0,\ n > 0$).

$$r_{xx}[m] = \begin{cases} r_{xx}^*[-m] & \text{for } m < 0 \\ -\sum_{n=1}^{p} a[n]r_{xx}[m-n] + \sigma_u^2 & \text{for } m = 0 \\ -\sum_{n=1}^{p} a[n]r_{xx}[m-n] & \text{for } m > 0 \end{cases} \tag{B14}$$

Using our results from eqn. (B14), we calculate the VAR parameters as follows

$$\begin{bmatrix} r_{xx}[0] & r_{xx}[-1] & \cdots & r_{xx}[-p] \\ r_{xx}[1] & r_{xx}[0] & \cdots & r_{xx}[-p+1] \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}[p] & r_{xx}[p-1] & \cdots & r_{xx}[0] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} = \begin{bmatrix} \sigma_u^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{B15}$$

Echoing the early discussion of VAR models, we have arrived at the AR Yule-Walker normal equations. It is also possible to calculate the power spectral density of an AR model from the autocorrelation sequence.

$$P_{AR}(f) = \frac{T\sigma_u^2}{|A(f)|^2} = T\sum_{k=-\infty}^{\infty} r_{xx}[k]e^{-i2\pi fkt} \tag{B16}$$

where T is the sampling frequency. Thus we can use the auto-correlation sequence to compute parameters of a vector autoregressive model.

[1] D. Weisburd, G. J. Bruinsma, and W. Bernasco, in *Putting Crime in its Place: Units of Analysis in Geographic Criminology*, edited by D. Weisburd, W. Bernasco, and G. J. Bruinsma (Springer, 2009).

[2] D. Oberwittler and P.-O. H. Wikstrm, in *Putting Crime in its Place: Units of Analysis in Geographic Criminology*, edited by D. Weisburd, W. Bernasco, and G. J. Bruinsma (Springer, 2009).

[3] E. Groff, D. Weisburd, and N. A. Morris, in *Putting Crime in its Place: Units of Analysis in Geographic Criminology*, edited by D. Weisburd, W. Bernasco, and G. J. Bruinsma (Springer, 2009).

[4] G. Lafree, Annual Review of Sociology **25**, 145 (1999), URL http://dx.doi.org/10.1146/annurev.soc.25.1.145.

[5] G. L. Kelling and J. Q. Wilson, The Atlantic (1982).

[6] R. J. Sampson and S. W. Raudenbush, Social Psychology Quarterly **67**, 319 (2004), ISSN 01902725, URL http://dx.doi.org/10.2307/3649091.

[7] R. J. Sampson, S. W. Raudenbush, and F. Earls, Science **277**, 918 (1997), URL http://dx.doi.org/10.1126/science.277.5328.918.

[8] J. D. Morenoff, R. J. Sampson, and S. W. Raudenbush, Criminology **39**, 517 (2001), ISSN 1745-9125, URL http://dx.doi.org/10.1111/j.1745-9125.2001.tb00932.x.

[9] L. J. Krivo and R. D. Peterson, American Sociological Review **65**, 547 (2000), ISSN 00031224, URL http://dx.doi.org/10.2307/2657382.

[10] M. D. Maltz, in *Putting Crime in its Place: Units of Analysis in Geographic Criminology*, edited by D. Weisburd, W. Bernasco, and G. J. Bruinsma (Springer, 2009).

[11] G. E. Tita and R. T. Greenbaum, in *Putting Crime in its Place: Units of Analysis in Geographic Criminology*, edited by D. Weisburd, W. Bernasco, and G. J. Bruinsma (Springer, 2009).

[12] C. W. J. Granger, Econometrica **37**, 424 (1969), ISSN 00129682, URL http://dx.doi.org/10.2307/1912791.

[13] C. W. J. Granger, pp. 48–70 (2001), URL http://portal.acm.org/citation.cfm?id=781843.

[14] H. Y. Toda and P. C. B. Phillips, Econometric Reviews **13**, 259 (1994), URL http://dx.doi.org/10.1080/07474939408800286.

[15] R. S. Hacker and A. Hatemi-J, Applied Economics **38**, 1489 (2006), ISSN 0003-6846, URL http://dx.doi.org/10.1080/00036840500405763.

[16] S. L. Marple, *Digital spectral analysis: with applications* (Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1986), ISBN 0-132-14149-3, URL http://portal.acm.org/citation.cfm?id=21429.

[17] A. Arnold, Y. Liu, and N. Abe, in *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, New York, NY, USA, 2007), pp. 66–75, ISBN 978-1-59593-609-7, URL http://dx.doi.org/10.1145/1281192.1281203.

[18] A. Schlogl, Signal Processing **86**, 2426 (2006), ISSN 01651684, URL http://dx.doi.org/10.1016/j.sigpro.2005.11.007.

[19] A. Neumaier and T. Schneider, ACM Trans. Math. Softw. **27**, 27 (2001), ISSN 0098-3500, URL http://dx.doi.org/10.1145/382043.382304.

[20] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis* (Springer, 2007), 1st ed., ISBN 3540262393, URL http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20\&path=ASIN/3540262393.

[21] J. Wooldridge, *Introductory Econometrics: A Modern Approach (with Economic Applications, Data Sets, Student Solutions Manual Printed Access Card)* (South-Western College Pub, 2008), 4th ed., ISBN 0324581629, URL http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20\&path=ASIN/0324581629.

[22] M. Tumminello, F. Lillo, and R. N. Mantegna (2008), 0809.4615, URL http://arxiv.org/abs/0809.4615.

[23] L. Laloux, P. Cizeau, J. P. Bouchaud, and M. Potters, Physical Review Letters **83**, 1467 (1999), URL http://dx.doi.org/10.1103/PhysRevLett.83.1467.

[24] K. B. K. Mayya and R. E. Amritkar (2006), cond-mat/0601279v1, URL http://arxiv.org/abs/cond-mat/0601279v1.

[25] A. Utsugi, K. Ino, and M. Oshikawa (2003), cond-mat/0312643, URL http://arxiv.org/abs/cond-mat/0312643.

[26] A. Edelman, SIAM J. Matrix Anal. Appl. **9**, 543 (1988), ISSN 0895-4798, URL http://dx.doi.org/10.1137/0609045.

[27] A. M. Sengupta and P. P. Mitra, Physical Review E **60**, 3389 (1999), URL http://dx.doi.org/10.1103/PhysRevE.60.3389.

[28] J.-P. Bouchaud, L. Laloux, M. A. Miceli, and M. Potters (2005), physics/0512090v1, URL http://arxiv.org/abs/physics/0512090v1.

[29] K. B. K. Mayya and M. S. Santhanam (2007), pp. 69–75, URL http://dx.doi.org/10.1007/978-88-470-0665-2_5.

[30] S. Thurner and C. Biely, Acta Physica Polonica B **38** (2007), URL http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=2007AcPPB..38.4111T.

[31] C. Biely and S. Thurner (2006), physics/0609053, URL http://arxiv.org/abs/physics/0609053.

[32] M. Winterhalder, B. Schelter, W. Hesse, K. Schwab, L. Leistritz, D. Klan, R. Bauer, J. Timmer, and H. Witte, Signal Process. **85**, 2137 (2005), ISSN 0165-1684, URL http://dx.doi.org/10.1016/j.sigpro.2005.07.011.

[33] M. Kamiński, M. Ding, W. A. Truccolo, and S. L. Bressler, Biological Cybernetics **85**, 145 (2001), URL http://dx.doi.org/10.1007/s004220000235.

[34] P. J. Franaszczuk and G. K. Bergey, Brain Topography **11**, 13 (1998), URL http://dx.doi.org/10.1023/A:1022262318579.

[35] L. A. Baccalá and K. Sameshima, Biological cybernetics **84**, 463 (2001), ISSN 0340-1200, URL http://view.ncbi.nlm.nih.gov/pubmed/11417058.